The Report committee for Joshua William Chrisman

Certifies that this is the approved version of the following report:

Studying for USMLE Step 1 using evidence-based strategies

SUPERVISING COMMITTEE:

Veronica Yan, Supervisor

Elizabeth Nelson

Studying for USMLE Step 1 using evidence-based strategies

by

Joshua William Chrisman

Report

Presented to the Faculty of the Graduate School

of the University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree(s) of

Doctor of Medicine

And

Master of Education

The University of Texas at Austin

May 2020

Studying for USMLE Step 1 using evidence-based strategies

by

Joshua William Chrisman, MD/MEd

The University of Texas at Austin, 2020

SUPERVISOR: Veronica Yan

The United States Medical Licensing Exam (USMLE) Step 1 is the first in a series of medical licensing exams that students take as they progress through medical school. It covers much of the basic science and foundational medical concepts covered during the first two years of pre-clinical education at most medical schools. Despite the facts that mastery of this knowledge is needed as a strong foundation on which to build the rest of a student's medical expertise and that scores on this test have a substantial impact on future training opportunities, there is virtually nothing in the medical education literature linking evidence based study strategies to Step 1 performance. In this paper, I review the literature on some of the most effective, evidence-based study approaches and lay out a study plan for students preparing to take the USMLE Step 1 exam.

Table of Contents

Motivation

During a medical student's years spent in training, there is arguably no more important test that is taken than the United States Medical Licensing Exam (USMLE) Step 1, more commonly referred to as USMLE Step 1 or simply, Step 1. This is the first in a series of three exams that must be completed and passed to be licensed as a physician in the United States (Gauer & Jackson, 2017). Step 1 is generally taken at the end of the first two years of medical school and covers an extensive base of basic science knowledge related to medicine (McDougle et al., 2013). The stakes surrounding Step 1 are high for medical students. Medical students that pass Step 1 on their first try are more likely to actually graduate from medical school (Andriole & Jeffe, 2012; *Graduation and Attrition Rates of U.S. Medical Students*, 2018) and about twice as likely to ultimately become specialty board certified compared to their peers that fail the exam one or more times (McDougle et al., 2013). Students' ability to obtain their desired residency is also affected by this exam. Step 1 performance affects decisions by residency programs on which students receive interviews (*Results of the 2016 NRMP Program Director Survey*, 2016) and which students match (*Charting Outcomes in the Match for U.S. Allopathic Seniors*, 2016).

There are major changes coming to the way this test is scored. Recently the USMLE announced a decision to change the scoring of the USMLE Step 1 exam from a three-digit numeric score to a pass/fail score (United States Medical Licensing Examination | Invitational Conference on USMLE Scoring, 2020). This clearly decreases the importance of maximizing performance on this exam as previously described. However, this does not mean that performing

well on this and other exams is not still important. Simply because there is no three-digit score attached to the exam does not mean the underlying factors that are driving differences in graduation rates and specialty certification will be diminished in any way. Seeing as Step 1 covers basic science concepts that underlie the pathophysiology, diagnosis, and treatment of many diseases, one could argue that one reason that students who have done poorly on this exam tend to have worse outcomes in their medical careers is that they have not built a strong set of foundational knowledge on which they can build the rest of their clinical knowledge and abilities. Cognitive load theory would suggest that the ability to perform the kind of complex and cognitively demanding tasks required of any physician can only be accomplished by freeing up working memory by having the foundational and basic component parts of the task committed to long term memory which is not subject to the kind of limitations of working memory (Sweller et al., 2019). The literature on development of medical expertise supports that experts in the field must be able to draw on a substantial and varied body of knowledge in order to create a mental model of the situation and construct possible (and correct) diagnoses (Lesgold et al., 1988). This knowledge includes foundational knowledge tested on Step 1 such as anatomy and physiology. Any medical student reading this can likely relate to the feeling that the ability to learn and recall all the medical knowledge required to diagnose and treat patients seems incredibly daunting at first. It is only through developing a strong understanding of the basic principles that the next level of knowledge seems attainable. This layering of ever more deep and complex medical

knowledge continues as the trainee advances through the decade or more that is required to fully complete medical training.

In addition to the need for building a strong base of clinical knowledge, Step 2 CK, the second part of the series of Step exams, will continue to receive a three-digit score (United States Medical Licensing Examination | Invitational Conference on USMLE Scoring, 2020). There is speculation that residency programs will simply transition to replacing their current reliance on Step 1 scores with use of Step 2 CK scores (USMLE Step 1 pass/fail winners and losers, 2020). Not only that, medical trainees will continue to face the challenge of taking multiple board exams throughout their training and medical careers. These include the Step exams, NBME subject exams during the core clerkship year, in-service exams during residency, and board certification exams after graduating residency. Although the importance of Step 1 will decrease in 2 years when the scoring changes take effect, the prevalence and importance of being able to effectively study for and pass these types of exams will not. Therefore, I believe that the content of this paper can still be applied to studying not only for USMLE Step 1 but to these other exams as well.

Although there is an extensive research base within the educational psychology literature, the strategies that this research supports are rarely intuitive and are not often explicitly taught to students in general, including medical students. It would be challenging and time consuming for any medical student to formulate a study plan based on the currently published evidence on their own, and most medical students likely are not even aware that this body of literature even exists.

Studying for USMLE Step 1 using evidence-based strategies

The goal of this paper is to review the medical education and educational psychology literature and compile a list of evidence-based study suggestions for use by medical students as they prepare for this, and the myriad other important examinations they will take throughout their medical careers.

Theoretical background:

Current Step 1 Literature. There is very little published information linking any specific study strategies to higher performance on USMLE Step 1. Much of the research published on the subject has explored the link between students' historical factors such as prior grades and MCAT scores to attempt to build predictive models (Burns & Garrett, 2015; Gandy, College, Herial, Khuder, & Metting, 2008; Giordano, Hutchinson, & Peppler, 2016; Gohara et al., 2011). However, this does nothing to inform students on how to actually approach the exam. There have also been a few studies that have looked at common resources used by students to prepare for the exam. Although, these only discussed materials used, not strategies on how to most effectively use these resources (Bonasso, Lucke-Wold, Reed, Bozek, & Cottrell, 2015; Schwartz, Lineberry, Park, Kamin, & Hyderi, 2018; Thadani, Swanson, & Galbraith, 2000). Often students rely on anecdotal evidence and advice from upperclassmen and internet message boards to create their study plans. There are nearly as many published papers recommending advice from upperclassmen as there are studies suggesting any concrete, evidence-based study approaches (Schwartz et al., 2018; Strowd & Lambros, 2010; Zhang, Rauchwarger, Toth, & O'Connell, 2004). Commercial preparation courses are also a commonly used preparation strategy. These

courses tend to be quite expensive and claim to provide superior preparation for the exam. However, several studies have found that students do no better on Step 1 after taking these courses, even after controlling for the fact that it is usually students that are performing at a lower standard academically that take these commercial courses (Bonasso et al., 2015; Thadani et al., 2000; Werner & Bull, 2003; Zhang et al., 2004). The amount of protected time that students spend studying exclusively for the exam immediately before taking it has also been the subject of some research. These studies have shown mixed results, with some showing that more weeks of study time and more hours studied per week are correlated with higher scores (Kumar et al., 2015; Thadani et al., 2000), while other studies have found no such relationship (Bonasso et al., 2015; Burk-Rafel, Santen, & Purkiss, 2017; Giordano et al., 2016). However, these studies made no mention of the methods used during this dedicated study period, let alone the quality of study strategies. I believe that these inconsistent data linking overall and weekly study time to test performance is an indicator that students need more help in understanding how to more effectively and efficiently use the study time they have. In this case, it seems more is not always better, but more efficient could be.

I was able to identify a small number of studies that looked at the implementation of one or more specific study strategies and correlations with Step 1 scores. One of the most commonly identified strategies was increased use of third-party test question banks. Several studies found a positive correlation between the number of practice test questions attempted and Step 1 score (Baños, Pepin, & Van Wagoner, 2018; Bonasso et al., 2015; Burk-Rafel et al., 2017; Kumar et

al., 2015). This is, of course, not surprising as self-testing as a study strategy is well supported by the educational psychology literature (Deng, Gluckstein, & Larsen, 2015; Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013; Larsen, Butler, & Roediger III, 2008; Roediger III, Putnam, & Smith, 2011). Beyond these papers focused on test bank usage, there are limited others. I did find one study that seemed to have promising methodology. These researchers used the Learning and Study Strategies Inventory (LASSI). Using other performance data such as MCAT scores, undergraduate GPA, and preclinical coursework grades along with LASSI scores, a model was constructed to predict Step 1 scores. From this, it emerged that among the ten LASSI subscales only the concentration subscale was a significant predictor of Step 1 exam scores (West, Kurz, Smith, & Graham, 2014). However, results of this study should be interpreted cautiously; this model was based on a relatively small (n=79) convenience sample from a single medical school. And while it does make sense to use a standardized tool to measure use of strategies, the LASSI is an outdated measure that was developed prior to much of our current understanding of how students learn most effectively. Additionally, the LASSI was administered during medical school orientation, almost two years prior to these students actually taking Step 1. This means that any development of new study strategies during medical school, and there were likely many, would not have been captured by this study.

Another recent study, also looking to build a predictive model for Step 1 scores, included several interesting variables. Those were early study, meaning whether or not a student spent time dedicated specifically to studying for Step 1 prior to any dedicated study time they might

have just prior to the test, and number of passes reading a Step 1 review book. Results showed that students who engaged in early study had significantly higher exam scores and that there was a positive correlation between the number of times going through the review book and exam scores (Burk-Rafel et al., 2017). Finally, the last article I could find showing evidence for a specific study strategy looked at the use of flashcards. This study found a benefit from using student generated, but not from commercially available flashcards (Deng et al., 2015).

A gap in the current literature exists in examining the impacts of effective and ineffective learning strategies on Step 1 performance. While there is a plethora of research available on these strategies in general, there is a surprisingly small amount of evidence linking these strategies directly to Step 1 and no publications laying out specific and comprehensive recommendations to help students improve their overall learning of medical school material and performance on the USMLE Step 1 exam.

So what strategies actually work?

Despite the fact that there are few studies demonstrating a link between specific approaches to study and outcomes on the USMLE Step 1 exam, there is a wealth of evidence available within the educational psychology literature to guide students in creating an approach to studying for this and the myriad other important board exams they will take throughout their careers.

Studying for USMLE Step 1 using evidence-based strategies

While there are many options a student could choose when it comes to the approach they will take to studying, there are a handful that have a consistently strong base of evidence showing their efficacy across multiple learning scenarios and student populations. Because of this, and in order to facilitate the usability of the study plan laid out in this paper I chose to focus on four study strategies that I believed would be best suited to improving student learning throughout the first two years of medical school and during dedicated time spent studying just prior to the exam. These are: retrieval practice, distributed practice, interleaving, and elaborative interrogation. In the following sections I will briefly review some of the literature supporting these strategies. All four of these strategies are part of a larger category known as desirable difficulties.

Desirable difficulty. Often learners associate the feeling that things are coming easily as a sign that real learning is taking place. However, the opposite is often true. It can be when learners are having trouble and having to wrestle with the material that the best understanding is developed, connections are made to other topics, and long-term retention is solidified (Bjork & Bjork, 2011; McDaniel & Butler, 2010). An experience of difficulty can lead to either an experience of success or an experience of failure. Regardless of outcome, however, the experience of difficulty can enhance learning if the learner is able to engage active, elaborative processes. In other words, it is not the outcome that matters, but the process. Both failure and success scenarios could provide educational benefits for the learner. It is important to recognize, however, that not all difficulties instilled into the learning process fall into this category of

desirable difficulty. Difficulty simply for the sake of difficulty is not beneficial. Kapur (2016) sets out a framework categorizing strategies as productive failures (e.g., desirable difficulties), unproductive failures (e.g., undesirable difficulties), productive successes (which may also be desirable difficulties), and unproductive successes. Unproductive successes are those tasks that a learner is able to accomplish but do not lead to any meaningful long-term learning.

Kapur and Bielaczyc (2012) demonstrated in a study of productive failure that, when combined with consolidation and review after the fact, failure in a complex task led to improved performance and generalizability of learning when compared to direct instruction. The students in the productive failure group of this experiment were asked to solve complex mathematical problems without help from the teacher. Their performance on a post-test was compared to those that had received direct instruction from the teacher. While the students in the productive failure group failed to come up with the correct solution on the initial task, they outperformed those that received direct instruction on the post-test. If we understand why this type of productive failure improved learning it can be used to improve studying for board exams. The reason learning was improved for the students in the productive failure group in Kapur and Bielaczyc's study has to do with the way they interacted with the problem and possible solutions to the problem. These students had to try out multiple different ways of not only solving the problem, but also of how to represent the problem in the first place. This allowed these students to better understand the constraints of various strategies and drew their attention to critical features of the targeted concept (Kapur & Bielaczyc, 2012).

So, what makes a difficulty desirable? It is when a difficulty draws the students' attention to these critical features and concepts and forces them to think critically about the problem and possible solutions. This is what leads to a learner experiencing either productive failure or productive success. Kapur (2016) as well as Bjork and Bjork (2011) would argue that it is not the success of solving a problem or retrieving a correct answer that leads to learning, but more importantly the way in which the learner engages with the material as she attempts to complete the task. There are numerous conditions that can help lead to an activity being productive or unproductive. One of the key components to productive failure is activation of prior knowledge. Though a student may not have all the knowledge necessary to solve the problem, it is in activating prior knowledge to arrive at possible answers that deep learning can take place despite the outcome in the short-term (Kapur, 2016). For example, approaching a challenging practice exam question can be an example of productive failure; when a student comes across a question that seems difficult, they should focus on applicable knowledge they do have in an attempt to come up with possible routes to the answer. This may not be successful, but it does require them to access and assess prior knowledge in a way that will be helpful. The post-failure consolidation phase is extremely important as well. This requires identifying gaps in knowledge as well as positive and flawed aspects of proposed solutions. It may be challenging and time consuming, but approaching learning in this way leads to greatly improved long-term retention and flexibility in applying learning to novel situations (Kapur & Bielaczyc, 2012).

In another paper presenting a framework for understanding desirable difficulties, McDaniel and Butler (2010) conclude that it is congruence between the difficult task and the eventual criteria against which the learner will be tested that determines whether a difficult learning task is desirable. This means that it is necessary to understand the challenges that a student will face in any given Step 1 exam question and choose appropriately difficult study techniques in order to stimulate students to practice those tasks while studying. In questions on the exam, students are expected to recognize a disease being described based on a limited amount of information provided in a clinical vignette. Sometimes the question goes only this far, but many questions require that students make further connections either to recall other information about the specific disease or to recall information about a related but different disease. More broadly, when it comes to enhancing learning for a career in medicine, it is important that foundational knowledge is fluently accessible. Therefore, in order for difficulties presented during studying to be desirable, they must push students to perform these same tasks. Ultimately, though incorporating desirable difficulties may make learning seem more challenging, effortful and slow, and may lead to increased mistakes during studying, it is *because* learners are engaging more deeply with the material that using these techniques improves long term learning.

Students should be aware of and vigilant to avoid study tasks that fall into the category of unproductive success. These strategies often give the illusion of learning through the short-term performance success that the learner experiences during the task itself. However, they do not

lead to meaningful long-term learning. Tasks that fall into this category are often the rote memorization type activities (Kapur, 2016). While some rote memorization comes with the territory of medical education and many medical students are tempted to memorize decontextualized facts during Step 1 study, this type of studying should be minimized.

The following sections provide specific examples of desirably difficult strategies that students can use to make their study time more effective. Before exploring these examples, however, imagine what a typical medical student might go through in studying for Step 1. A typical medical student is given several weeks just prior to the exam in which they study 10+ hours a day, 7 days a week, cramming as much as possible before they sit for the test. Prior to this dedicated study period they may have done little to no studying specifically for the test, although the material on the test is mostly covered in the pre-clinical curriculum of medical school. During this brief study time medical students often read and reread several popular review books, watch review videos repeatedly, use pre-made flashcards aimed at rote memorization of medical facts, and go through thousands of practice questions. Some of these strategies can be effective, but as we will see in the following sections, there is also room for significant improvement.

Retrieval practice. In the example just given of a typical student's approach to the exam there is already some use of retrieval practice. The use of question banks to test knowledge during dedicated study time is nearly ubiquitous. However, practice exam questions are often used as a way to assess progress in studying rather than as a tool for learning itself. A more

evidence-based approach would be to start using practice exams much earlier as another way to learn the material. Additionally, instead of using pre-made flashcards aimed at rote memorization, a student could make their own cards that engage retrieval in a deeper and more meaningful way. The one thing to be careful about is to make sure that the types of questions (i.e., the type and level of retrieval that the student is engaging with) matches what students expect on the exam. These suggestions will be covered in more detail in this section and in the study plan presented later in the paper.

The practice of self-testing is one of the most reliable and empirically-robust desirable difficulties, with numerous studies demonstrating the effectiveness of using testing--not as an assessment tool--but as a pedagogical tool (Butler & Roediger, 2007; Dunlosky et al., 2013; Larsen et al., 2008; Roediger III et al., 2011). Retrieval practice, or learning through testing, has been shown to strengthen memory and improve recall of information. In fact, testing of previously learned information is a more powerful learning event than restudying that same material (Butler & Roediger, 2007; Karpicke, Lehman, & Aue, 2014; Larsen et al., 2009; Schmidmaier et al., 2011; Weinstein, Madan, & Sumeracki, 2018). This makes time spent self-testing previously learned information an immensely powerful strategy, since it acts as both learning by retrieval and distributed practice. Certain types of retrieval practice can have additional benefits. For example, the use of well written multiple-choice test questions, can improve learning and recall of not only the information applicable to the correct answer but also improve learning of information related to the wrong answer choices as well. This is because the

learner must recall this incorrect answer associated information as they eliminate answer choices (Little, Bjork, Bjork, & Angello, 2012). While there is some conflicting evidence about whether or not there is a benefit from testing using the multiple choice technique (Butler & Roediger, 2007), this discrepancy can be explained by the nature and quality of multiple choice tests used (Little et al., 2012), with context-rich multiple choice questions such as the type used on medical licensing exams and practice exams being the most beneficial (McConnell, St-Onge, & Young, 2015).

Distributed Practice. A student preparing for an exam that is one week away could study for seven hours the day before the exam, cramming in as much as they can just before the test. They could also choose to spend the same amount of time studying but spend just one hour per day each day for the entire week. This is a simple example of the contrasting approaches of massed versus distributed practice. Research has shown that the student that chooses to spread out their study time will experience much better long-term learning than the student that crams all their study together.

The spacing effect is the phenomenon in which long term retention is substantially increased by breaking the study of a subject into smaller time intervals and spreading those time intervals out over days or weeks (Dempster, 1989). This is in contrast to massed practice, studying a subject in one long session without returning to it again before the exam. The distributed practice option has been repeatedly shown to have a powerful impact on improving recall and long-term retention of material (Benjamin & Tullis, 2010; Kerfoot, DeWolf, Masser,

Church, & Federman, 2007; Matos, Petri, Mukamal, & Vanka, 2017; Pavlik & Anderson, 2005; Rohrer & Pashler, 2007). It has also been shown to improve the acquisition of psychomotor skills in the setting of surgical resident education (Moulton et al., 2006; Van Dongen et al., 2011).

The reason spacing presentations of material significantly improves the long-term retention of that material is actually related to retrieval practice. When material is seen repetitively in short succession, very little effort is required to retrieve that same material from memory. As the time since the last presentation increases, the effort required for retrieval is increased. It is theorized that this effort put forth to remember the information actually helps more deeply encode that information (Benjamin & Tullis, 2010; Pavlik & Anderson, 2005). This theory of the spacing effect helps explain why benefits of expanded distribution are also observed. A number of studies have shown that sequentially increasing the time between presentations of materials or retrieval practice sessions leads to improved long-term retention of that material (Dobson, 2011, 2012). This suggests that as material is better learned, a longer period of distribution is required to gain the benefit of distributing practice. In practical terms, this means that when material is brand new it should be covered more frequently, but over time that material can be returned to less and less frequently. Luckily for students, if they choose to use an electronic flashcard program, many of these have this function built in.

Incorporating distributed practice into studying has the advantage of providing substantial gains in long-term learning without increasing the actual amount of time spent studying (Rohrer

& Pashler, 2007), a fact that any busy medical student should be more than happy to take advantage of.

Interleaving. While it does not have quite the same widespread reputation as distributed practice and retrieval practice, interleaving is another strategy that has consistently been shown to improve student learning. Like distributed practice, interleaving is often described as an alternative to massed study. Interleaving can be used when a student needs to study several different topics. This is, of course, the case for students in medical school who are studying a staggeringly large number of topics and subtopics. Rather than studying all the information in a topic at once, interleaving involves breaking that studying into smaller chunks then intermixing those with study of other topics. Interleaving has been shown to be particularly good helping with inductive learning, or learning from examples (Kornell & Bjork, 2008). Fortunately for medical students, this type of learning is ubiquitous in medicine. Doctors in training must learn to diagnose a disease through their experiences diagnosing that disease in other similar patients. Not only that, but interleaving has been shown to help learners distinguish between similar but different concepts (Rohrer, 2012), which is a common problem on medical licensing exams. There are commonly diseases that present very similarly, and the test taker must pick up on subtle differences in order to correctly answer the question. In addition to these specific advantages, like distributed practice, interleaving has also been shown to improve general recall and long-term retention (Dunlosky et al., 2013; Weinstein et al., 2018) while requiring no additional time, simply restructuring of study time already being used.

Recently, Brunmair and Richter (2019) published a meta-analysis investigating the effect size of interleaving as well as how the effects of interleaving differ based on the type of material being studied. This sheds some interesting light on the topic and also suggests that there are times that interleaving should be used and other times it may not be as helpful. Interleaving was shown to be particularly helpful when dealing with visual material. This means that interleaving should be applied to topics such as anatomy that are highly visual and other topics that are frequently represented in visual material such as biochemical pathways or physiology. The overall theme that came out of the meta-analysis was that for these specific types of materials, interleaving is useful for helping students learn to discriminate between similar items in different categories. A student could, for example, use interleaving during a study session to study diseases that all present with muscle weakness. This would help the student be able to differentiate these diseases on test day.

There are, however, times when using a blocked approach (as opposed to interleaving is helpful). Specifically, when the goal is to be able to differentiate between examples within the same category and when recognizing the similarities that tie examples together within a category are crucial (Brunmair & Richter, 2019). It is those times that a group of diseases being studied appear very different but learning to recognize them as a group may be very important that blocking could be useful.

Elaborative Interrogation. An alternative to rereading a text multiple times in hopes of getting more out of it on subsequent passes is to instead ask and answer how and why questions

about the material periodically during the first reading of a text. This technique is a relatively simple one that can be integrated into almost any studying already being performed and has been shown to have a significant impact on learning. While this does take substantially longer to work through a text in this way while asking and answering these why questions, it has been shown to have rather robust effects on comprehension and recall.

The leading theory as to why elaborative interrogation is effective is that it encourages organization and structuring of information in memory and encourages activation of and making connections to existing knowledge (McDaniel & Donnelly, 1996; Weinstein et al., 2018), both of which aid in comprehension and retention. It also helps in the acquisition of deeper learning (Hattie & Donoghue, 2016). These improvements in comprehension and retention have been demonstrated specifically in experiments using challenging and complex scientific materials (McDaniel & Donnelly, 1996; Smith et al., 2010).

What doesn't work?

Perhaps as important as identifying and using those study strategies that do work well, is also the elimination of strategies that have been shown to be ineffective. Some of these have been mentioned previously throughout this paper, but they bear repeating as time spent on such strategies would yield much better results if used in other ways. Some commonly used, but low yield techniques are: rereading, highlighting, and summarizing (Dunlosky et al., 2013).

Rereading. This is a very commonly used approach amongst medical students. Especially during dedicated study time students will often go over books or old notes over and over again in hopes of drilling the information into their brains. This is not effective and should be avoided. That is, however, not to say that material should never be reread. This should be done strategically though, with lots of opportunities for retrieval and elaboration throughout. For example, a student, while reviewing a practice exam, will usually find that there is key information or concepts that they did not know or did not understand. This would be a good time to go back and address specific deficiencies by reviewing material.

Summarizing. This is another strategy that gets used often, likely because it does seem to have some face value appeal. Summarizing is trying to write out the key points of a text or lecture in a shortened written version. Some studies have shown summarization to be moderately useful, but only for students that have had some extensive training in how to properly summarize in a way that maximizes learning (Dunlosky et al., 2013). However, the time it would take to train students to summarize effectively is likely not worth the small benefit students might receive from such training; it is much more efficient to engage in retrieval practice.

Highlighting. While it may seem logical that identifying and highlighting key information in a text as one studies would improve comprehension and retention, there is not evidence to support this having substantial positive impact on learning. Similar to summarizing, there may be a small benefit to highlighting when done properly, but for most students there is no benefit, and could even hinder learning at times (Dunlosky et al., 2013).

Methods

I reviewed the educational psychology and medical education literature for effective and ineffective strategies linked to performance in high-stakes testing, then evaluated the evidence found in the literature to make recommendations for a Step 1 study plan.

As it would be nearly impossible to review all the current medical education and educational psychology literature to identify all possible effective study techniques. I initially relied on several previously published review articles that had identified particularly high yield study strategies (Dunlosky et al., 2013; Hattie & Donoghue, 2016; Rohrer & Pashler, 2010; Sumeracki & Weinstein, 2018; Weinstein et al., 2018). After identifying the target strategies of desirable difficulties, spaced practice, interleaving, retrieval practice, and elaborative interrogation based on these resources, I searched the literature via PUBMED, MEDLINE, ERIC, and PSYCHINFO databases for additional articles on how and why these strategies work for developing long term understanding, retention, and performance on exams.

Suggested Study Plan

The proposed USMLE Step 1 study plan can be broken down into two parts based on timeline, that is, a long-term portion that consists of Step 1 specific studying performed throughout the first two years of medical school, and the shorter dedicated study time that nearly all medical students utilize immediately prior to taking the exam. Early study should happen throughout the first two years of medical school, while I would suggest between four and six

weeks of dedicated study time immediately prior to the test. Inclusion of suggestions for studying specifically for Step 1 prior to this dedicated study time are based on evidence that this is associated with higher scores (Burk-Rafel et al., 2017) and the evidence that such spaced studying leads to better long-term retention and understanding both in general (Benjamin & Tullis, 2010; Dunlosky et al., 2013) and specifically within medical education (Kerfoot et al., 2007; Matos et al., 2017). It is likely that as scoring changes are implemented within the next couple of years that schools will no longer provide time off for students to dedicate time purely to studying for Step 1. In that case, the early, distributed study time suggested in this study plan would be the only studying a student would have prior to the exam.

Appendix A contains a layout of proposed study schedules for both long-term study during the preclinical years as well as during time given specifically to study prior to the exam.

Long Term Study

Considering the substantial evidence in favor of spaced practice. It will be important for students to begin studying for Step 1 early in their medical school experience. This will allow for better long-term retention and better integration of learning with other concepts. The following are the specific ways I suggest students go about studying Step 1 specific material during the first two years of medical school.

Resources to use early and how to use them during first year. Since learning necessarily progresses from surface level knowledge toward deeper knowledge and transfer (Hattie & Donoghue, 2016), the early stages of learning are the times to cover basic knowledge

and facts. Initially I suggest that students cover these basic topics using video resources such as Boards and Beyond (See Appendix B for a list of suggested resources that can be used and will be referenced throughout this study plan). For initial encoding of information, these types of videos, including narrated animations, are particularly effective (Mayer, 2008; Mayer & Anderson, 1992). I would suggest spending about 30 minutes to one hour watching videos from a resource such as Boards and Beyond every other day. Most of these videos are between 10 and 30 minutes in length, so this would represent between 2-6 videos per day. At this rate, a student should be able to review the entire Boards and Beyond library of videos during their first year. In order to incorporate interleaving, each video should change topic or subtopic area. On days that students are not studying from Boards and Beyond they should be incorporating retrieval practice by using a student generated flash card program such as Anki or Quizlet. As mentioned previously, student generated flashcards are more efficacious than premade flashcards (Deng et al., 2015). Students should, again, spend approximately 30 minutes to an hour on this task. They should split this time so that they spend about half generating new flash cards based on material recently studied on Boards and Beyond, or whichever resource they are choosing to use, and the other half reviewing past flashcards. Generating new flashcards will be extremely helpful in learning material for several reasons. First, it requires students to make judgments about what they have learned well and focus on those areas that they believe are weak. Creating flashcards the day after seeing material forces students to make judgements about their comprehension following a delay after seeing the material. This has been shown to significantly improve

metacomprehension, especially when the task is a generative one (Thiede, Anderson, & Therriault, 2003; Thiede, Dunlosky, Griffin, & Wiley, 2005). In order to make this process more generative and to encourage deeper encoding of information, students should focus on making flashcards that make connections to other material they have learned as well. This is also a chance to intentionally incorporate elaborative interrogation, asking how and why concepts are related. There are, of course, times when flashcards will be much more purely memorization, but that is okay, especially in this early phase of learning. Part of the point of this initial review of material is to simply have exposure to the material. Reviewing flashcards is clearly an example of retrieval practice, the evidence for which has been discussed previously.

While reviewing flashcards can be an excellent tool to stimulate long term learning through retrieval practice, it is important for students to design and use flashcards in a way that will require them to retrieve information from memory and make connections between different topics. As with all the strategies suggested for studying that will be described in this paper, the amount that a learner will get out of any particular modality of studying depends on the way in which they engage with the material and requires a thoughtful and effortful approach. Use of flashcards is no different. Some students create, or even worse, buy flashcards that simply have facts recorded on them. They then flip through these cards reading them. This is no better than reading and rereading a text and is not an efficient use of time (Dunlosky et al., 2013). At a minimum, flashcards should have one side that asks a question or has a prompt requiring the student to generate some kind of response. The other side of the card should have the answer.

However, the student must come up with the answer before looking at the back of the card. Otherwise they are not getting the benefit of retrieval practice. An example of the most basic type of card would be as follows:

Side 1: On which Chromosome is the gene involved in the development of Huntington Disease?

Side 2: Chromosome 4

Now this, of course, only requires the student to retrieve a single fact abstracted from all other knowledge of the disease and related diseases. A more in-depth card that would result in more retrieval and learning might look like the following:

Side 1: Name three neurodegenerative diseases and describe how they can be differentiated based on changes in neurotransmitter levels.

Side 2: Huntington Disease – decreased acetylcholine and GABA, increased dopamine

Parkinson Disease – increased acetylcholine and decreased dopamine

Alzheimer Disease – decrease acetylcholine

This type of card requires the learner to retrieve much more information and to compare the various diseases based on key attributes, a task that will result in much better learning than the simpler example. Other more in-depth cards might ask the student to draw out a biochemical

pathway, list as many diseases as they can think of that have a genetic basis on the same chromosome or compare and contrast two similar diseases.

Students could also try creating activity cards that have them do a more in-depth dive into a particular topic. For example, they could create one deck of cards, each listing a commonly tested disease, and a second deck with activity cards such as “draw out the biochemical pathway involved in this disease”, “describe the basic pathophysiology of this disease”, or “describe the clinical presentation of this disease including signs, symptoms, lab and imaging abnormalities.” They could then spend some of their flashcard time drawing a card from each deck at random and performing the prescribed activity.

As was discussed earlier, it is important that study practices match the exam requirements in that studying should simulate the same types of processes that will be required in answering test questions (McDaniel & Butler, 2010). It is important for students to understand the structure of the exam and questions in order to create activities and flashcards that simulate the same processing demands. The questions on Step 1 and other similar board exams range from simple recall questions in which the test taker is asked a very direct question expecting them to simply recall a fact such as “What is the most common pathogen causing pneumonia in patients with cystic fibrosis?” It can be as simple as that, either you know it or you don’t. This is why it is necessary to memorize a lot of facts for these exams. However, compared to other types of questions, these purely recall based questions are somewhat rare. Students are more likely to encounter more complicated questions like those that give a lengthy clinical vignette describing a

patient's presentation, often accompanied by lab values and/or imaging. The question would then ask something such as "What is the most likely diagnosis?" or "Based on the most likely diagnosis, what is the genetic abnormality?" The test taker must be able to recognize, based on a likely incomplete list of signs, symptoms, lab values, and imaging, which disease is being described. This requires the ability to not only recognize multiple data points that suggest the correct diagnosis, but also to potentially remember attributes of similar, but incorrect diagnoses in order to narrow it down to the right answer. In the first example question, the names of the disease would be given as answer choices and therefore may act as memory aids, but in the second example only a list of genetic abnormalities would be given as choices, requiring the student to also recall which disease goes with each genetic abnormality. Creating flashcards/activities that expect these same processes from students will be helpful in improving performance on test day. Students should be able to recognize a disease based on being shown a presentation, but also need to be able to recall other associated information about that disease. They also need to be able to compare and contrast similar diseases so that they can differentiate them.

Students should also begin testing themselves early using a question bank of practice USMLE style questions. I would recommend using a question bank other than UWorld such as Kaplan or USMLE Rx (UWorld should be saved for dedicated study time). Some students may feel very challenged by this suggestion and may be troubled by the likely scenario in which they will be getting a large number of questions incorrect. This is normal and expected; students need

to realize that failure is not only acceptable, it is a positive thing. There is evidence that thinking through challenging questions, even if one comes up with an incorrect answer, has significant benefits for long term learning (Yan, Clark, & Bjork, 2016). Students should avoid the trap of only engaging in activities that make them feel like they are learning easily, because this is not always congruent with actual, long-term learning. Some learning activities will feel very difficult and result in failure, sometimes these will be some of the most beneficial (Kapur, 2016), especially if they involve similar forms of processing that will be required on the exam (McDaniel & Butler, 2010). In this case, that would be true, working through difficult practice questions is exactly the same type of processing that students will be asked to engage in on the actual Step 1 exam.

I suggest that students go through a small number of questions each day. A student that does just six questions a day could complete the entirety of the Kaplan question bank during their first year of medical school. During this first year, questions should be selected only from content areas that students have covered. For example, if the first course of the medical school is biochemistry, questions should be taken from this area only until other topics have been covered. Additional topics can be added as those courses are covered. This doesn't mean that students will not be exposed to new information while doing these questions, but they will at least have some background knowledge and frame of reference. Attempting to answer questions from topics to which students have no exposure would likely result in unproductive failure. Doing five or six questions in ten minutes, then spending another ten minutes reviewing the answers should be

sufficient. If this were combined with 40 minutes of Boards and Beyond review/Anki flashcards, these tasks would require one hour per day. First year medical students should be able to find time to fit this into their schedules, especially since much of this content will overlap with what is being learned in classes and therefore this study time will be adding to, not taking away from studying for medical school classes.

Study time during second year. Planned study time during this year should look very similar to the time spent during first year, just with some minor adjustments. These adjustments will focus on increasing the amount of self-testing and providing additional distributed practice by reviewing most of the same topics covered during first year, but through different sources. This will be the time for students to be studying and learning from the resource *First Aid for the USMLE Step 1*, one of the most popular resources used by students. However, in my opinion this is also one of the most misused resources. This book does a good job of presenting a lot of surface level facts in the form of bulleted lists and tables. Many students spend a great deal of time reading, rereading, and highlighting this book, all strategies that are not very efficacious (Dunlosky et al., 2013). I do think that there is value in covering this material, it just needs to be done in an effective manner. I would suggest reviewing this book throughout second year while utilizing both interleaving and distributed practice. Students should spend approximately 30 minutes every third day reviewing from this book. Some of this time should be spent reviewing past material and some should be spent studying new material. Students should be spending only 10 to 15 minutes on any one topic then switch to a new topic as to incorporate interleaving into

their studying. Days spent studying *First Aid* should be every third day. It is likely that students will finish *First Aid* prior to the end of the year. That is okay, and students should resist the possible urge to go back and start the book over again. At this point they've already seen everything in it and hopefully incorporated much of the salient information into their student generated flashcards. Students that find themselves in a position where they have finished *First Aid* before the end of the year should switch to another high-yield resource for this study time. We would suggest *Pathoma*, another of the most popular study resources for Step 1. Students could incorporate both the video lectures and textbook into this study time. Working with student created Anki flashcards should also continue in the exact same manner as first year; the number of days will decrease slightly since days spent using Anki should be the day after days spent studying *First Aid*.

Self-testing should be increased during second year. Now that students have had at least one pass over most of the material, retrieval practice will become one of their most important tools for solidifying this content, deepening their understanding, and ensuring that they can apply it usefully on test day. By studying *First Aid* only every third day, one day out of every three will be open to spend the entire hour of Step 1 study time doing practice questions. During second year, all practice questions should be selected at random from the question bank rather than from specific topic areas. Since students have now seen most, if not all of the material, the concern over undesirable difficulty no longer applies. By doing randomly selected questions, students will be incorporating interleaving into their question time in addition to retrieval practice. This

should take the form of doing 10 questions, then spending the rest of the time reviewing the answers. This is a very important part of the process. Spending time reviewing and asking oneself why the right answer is right and the wrong answers are wrong and how they thought through the process and came up with the answer will employ elaborative interrogation. With increasing the number of questions done on this test taking day, students should be able to easily get through another 2000+ questions during second year. They should do this without using UWorld. If they have followed the study plan up to this point, they will have completed somewhere around 4000 unique practice questions. This should already put students in an excellent place to be successful; there is a strong correlation between number of unique question bank items attempted and final Step 1 performance (Baños et al., 2018; Burk-Rafel et al., 2017).

Studying during time dedicated just to studying prior to the exam

First of all, while not explicitly covered in this current paper, it bears acknowledging that this period of time can be very taxing mentally, emotionally, and physically for students. It is important to remember that preparing themselves mentally and motivationally are equally important to their academic preparedness. Students should engage actively in self-care and prepare themselves mentally and emotionally for a grueling study period and test day.

Additionally, it is highly likely that schools will dispense with giving students weeks to months to study for Step 1 once the exam has been transitioned to a pass/fail scoring system. In that case, the following study plan would obviously not be implemented. That being said, the point of the dedicated study period is extremely performance oriented. In terms of long-term

learning there is likely little students will take away from this time, as most students who have been through this can likely attest. It is one of the most extreme examples of cramming for an exam and is aimed at maximizing performance in the short term. If a student has followed the previously suggested study strategies throughout their pre-clinical years, they should be more than prepared to take and pass the exam and should retain useful information that will serve them well as their medical training continues.

For most medical students, this period of dedicated study time will last between four and eight weeks immediately following the completion of the second year of medical school and immediately preceding administration of Step 1. This time will be spent focused heavily on preparing via testing. Students should aim to get through the entirety of the UWorld question bank during this time. The UWorld question bank most closely resembles the real test and there is evidence showing that this question bank is associated with the greatest increases in student scores (Bonasso et al., 2015). All questions should be done in 40 question blocks set to randomly draw questions from the question bank. This most closely simulates the real test and will ensure that students are incorporating interleaving into their study. In order to finish UWorld in this time period, students will need to do two of these blocks six days a week (if they are taking five weeks of study). This will obviously be adjusted based on the length of the dedicated study period.

Review of these questions is as important as the results. As discussed earlier, getting questions wrong is expected and not a bad thing. The important part is that students are learning

from this testing. Again, they should focus on spending time explaining to themselves why the right answer is right, why the wrong answers were wrong, and how they came to the answer they chose. This will engage students in elaborative interrogation and improve their understanding of the material.

This testing and review will take up approximately half a day of studying. Students should schedule this as best fits their motivation and schedule. The other half of the day should be spent reviewing material in areas of weakness. These can be identified through previous tests, CBSE exams, and the UWorld question bank app. The idea here is that students will be filling in their gaps in knowledge. They should choose these topics then spend time reviewing them in an interleaved manner. Students can use a variety of resources as well for this review, e.g. old lecture notes, review books, review videos, etc… But they should make sure that they are spending brief periods of time on each topic, then switching to another.

This study strategy as described for dedicated time incorporates spaced repetition, interleaving, retrieval practice, and elaborative interrogation. There should still be plenty of challenges in the UWorld questions, if there isn't and a student finds every question to be easy, they are ready and should just go take the test. There is a strong positive correlation between percent correct on UWorld questions and actual Step 1 performance (Giordano et al., 2016).

Students may also want to take several Comprehensive Basic Science Self Assessments through the National Board of Medical Examiners. These self-assessments have been shown to have a strong positive correlation with actual Step 1 performance (Morrison et al., 2010). Taking

these could provide helpful prognostic and diagnostic information for students throughout their study time.

While some of this advice for dedicated study time may seem vague, I believe that this is necessary. The goal of the spaced study time during the first two years was to build a solid, long-term base of knowledge and understanding of concepts. Now that students have done that, this is the time for them to personalize their study to their own needs and self-identified gaps in knowledge.

Limitations

Some of the resources suggested in this study plan are not free. Boards and Beyond, *First Aid for the USMLE Step 1,* Kaplan QBank, and UWorld Qbank all cost money, some of these cost hundreds of dollars. It must be acknowledged that this may be a burden on some students. However, I also recognize that most students are already spending money on these resources. Often students are spending much more on commercial prep courses. There are likely free alternatives to these resources that exist, and a future iteration of this plan may suggest alternative resources to use in the place of those that must be paid for. The rationale for using the resources I chose was to incorporate the very best resources I could find. Wherever possible this distinction was made based on research (Bonasso et al., 2015).

Conclusion

There is currently a lack of evidence-based study strategies pertaining to Step 1 in the medical education literature. Despite lacking evidence of the utility re-reading (Dunlosky et al., 2013), many students report re-reading a review textbook up to three times as a common study strategy. This paper presents a study plan grounded in current educational psychology literature that eliminates the use of ineffective strategies so students can use their time more efficiently. This study plan discourages the use of such ineffective strategies as highlighting, summarization, and rereading (Dunlosky et al., 2013). It includes the use of highly effective strategies such as spaced practice (Benjamin & Tullis, 2010), interleaving (Rohrer, 2012), and retrieval practice (Larsen et al., 2008; Roediger III et al., 2011). Use of a plan such as the one presented here should help students develop a better foundational set of knowledge they can draw on for the rest of their medical training and career and improve their performance on the exam.

Tables

Example weekly schedules throughout the study period							
Year 1							
	Monday	**Tuesday**	**Wednesday**	**Thursday**	**Friday**	**Saturday**	**Sunday**
Activity	Watch 2-6 Boards and Beyond Videos (number will vary based on length) + 5-6 Qbank Questions and Review	Flashcard creation and review previously made flashcards + 5-6 Qbank Questions and Review	Watch 2-6 Boards and Beyond Videos (number will vary based on length) + 5-6 Qbank Questions and Review	Flashcard creation and review previously made flashcards + 5-6 Qbank Questions and Review	Watch 2-6 Boards and Beyond Videos (number will vary based on length) + 5-6 Qbank Questions and Review	Flashcard creation and review previously made flashcards + 5-6 Qbank Questions and Review	Day off
Time required	1 - 1.5 hours	1 - 1.5 hours	1 - 1.5 hours	1 - 1.5 hours	1 - 1.5 hours	1 - 1.5 hours	0 Hours

Strategies Utilized	Interleaving, Distributed Practice, Retrieval Practice	Interleaving, Distributed Practice, Retrieval Practice, Elaborative Interrogation	Interleaving, Distributed Practice, Retrieval Practice	Interleaving, Distributed Practice, Retrieval Practice, Elaborative Interrogation	Interleaving, Distributed Practice, Retrieval Practice	Interleaving, Distributed Practice, Retrieval Practice, Elaborative Interrogation	
Other Details	Choose topics from those related to blocks already covered or currently being taught in curriculum. This applies to videos and questions.	Create new flashcards based on material covered on Monday in video review. **Important note:** *see body of paper for more details about how to do this effectively*	Choose topics from those related to blocks already covered or currently being taught in curriculum	Create new flashcards based on material covered on Wednesday in video review	Choose topics from those related to blocks already covered or currently being taught in curriculum	Create new flashcards based on material covered on Friday in video review	It is important to consider your personal life and maintaining balance as well as mental/physical health, take the day off studying. This doesn't have to be Sunday, put this day where it makes the most sense for your schedule

Table 1 - Example weekly schedule during first year study period

Studying for USMLE Step 1 using evidence-based strategies

Year 2							
	Monday	**Tuesday**	**Wednesday**	**Thursday**	**Friday**	**Saturday**	**Sunday**
Activity	Book review from either First Aid or Pathoma + 5-6 Qbank Questions and Review	Flashcard creation and review previously made flashcards + 5-6 Qbank Questions and Review	Answer and review 10-15 Qbank questions	Book review from either First Aid or Pathoma + 5-6 Qbank Questions and Review	Flashcard creation and review previously made flashcards + 5-6 Qbank Questions and Review	Answer and review 10-15 Qbank questions	Day Off
Time required	1 - 1.5 hours	1 - 1.5 hours	1 - 1.5 hours	1 - 1.5 hours	1 - 1.5 hours	1 - 1.5 hours	0 Hours

Strategies Utilized	Interleaving, Distributed Practice, Retrieval Practice	Interleaving, Distributed Practice, Retrieval Practice, Elaborative Interrogation	Interleaving, Distributed Practice, Retrieval Practice, Elaborative Interrogation	Interleaving, Distributed Practice, Retrieval Practice	Interleaving, Distributed Practice, Retrieval Practice, Elaborative Interrogation	Interleaving, Distributed Practice, Retrieval Practice, Elaborative Interrogation	
Other Details	Interleave topics. Spend no more than 10-15 minutes on any one book topic or subtopic.	Create new flashcards based on material covered on Monday First Aid/Pathoma review	Thoroughly review correct and incorrect answers. Time spent reviewing answers can also be a good time to make new flashcards	Interleave topics. Spend no more than 10-15 minutes on any one book topic or subtopic.	Create new flashcards based on material covered on Thursday First Aid/Pathoma review	Thoroughly review correct and incorrect answers. Time spent reviewing answers can also be a good time to make new flashcards	Again, doesn't have to be Sunday. Put this where it makes the most sense for you.

Table 2 - Example study schedule during second year study period

Dedicated Study							
	Monday	**Tuesday**	**Wednesday**	**Thursday**	**Friday**	**Saturday**	**Sunday**
Activity	Practice Exam + Review of answers	2 x 40 Uworld practice questions answered and reviewed + Review areas of weakness and flashcards	2 x 40 Uworld practice questions answered and reviewed + Review areas of weakness and flashcards	2 x 40 Uworld practice questions answered and reviewed + Review areas of weakness and flashcards	2 x 40 Uworld practice questions answered and reviewed + Review areas of weakness and flashcards	2 x 40 Uworld practice questions answered and reviewed + Review areas of weakness and flashcards	Day Off
Time required	8 - 10 hours	8 - 10 hours	8 - 10 hours	8 - 10 hours	8 - 10 hours	8 - 10 hours	0 Hours

Table 3 - Example study schedule during dedicated study period

Studying for USMLE Step 1 using evidence-based strategies

Suggested Study Resources	
Question Banks	UWorld Kaplan USMLE Rx Amboss
Video Libraries	Boards and Beyond Pathoma Sketchy (Micro, Pharm, Path)
Review Books	First Aid for USMLE Step 1 Pathoma
Other	Medbullets - step1.medbullets.com (great short summarization of many topics, excellent to be used as an additional review source when looking up information you may have missed; also it's free)

Table 4 - Suggested Study Resources

References

Andriole, D. A., & Jeffe, D. B. (2012). A National Cohort Study of U.S. Medical School Students Who Initially Failed Step 1 of the United States Medical Licensing Examination: *Academic Medicine*, *87*(4), 529–536. https://doi.org/10.1097/ACM.0b013e318248dd9c

Baños, J. H., Pepin, M. E., & Van Wagoner, N. (2018). Class-Wide Access to a Commercial Step 1 Question Bank During Preclinical Organ-Based Modules: A Pilot Project. *Academic Medicine*, *93*(3), 486–490. https://doi.org/10.1097/ACM.0000000000001861

Benjamin, A. S., & Tullis, J. (2010). What makes distributed practice effective? *Cognitive Psychology*, *61*(3), 228–247. https://doi.org/10.1016/j.cogpsych.2010.05.004

Bjork, E. L., & Bjork, R. A. (2011). Making Things Hard on Yourself, But in a Good Way: Creating Desirable Difficulties to Enhance Learning. In *Psychology and the real world: Essays illustrating fundamental contributions to society.* (p. 10). Worth Publishers.

Bonasso, P., Lucke-Wold, B., Reed, Z., Bozek, J., & Cottrell, S. (2015). Investigating the Impact of Preparation Strategies on USMLE Step 1 Performance. *MedEdPublish*. https://doi.org/10.15694/mep.2015.004.0005

Brunmair, M., & Richter, T. (2019). Similarity matters: A meta-analysis of interleaved learning and its moderators. *Psychological Bulletin*, *145*(11), 1029–1052. https://doi.org/10.1037/bul0000209

Burk-Rafel, J., Santen, S. A., & Purkiss, J. (2017). Study Behaviors and USMLE Step 1 Performance: Implications of a Student Self-Directed Parallel Curriculum. *Academic Medicine*, *92*, S67–S74. https://doi.org/10.1097/ACM.0000000000001916

Burns, E. R., & Garrett, J. (2015). Student failures on first-year medical basic science courses and the USMLE step 1: A retrospective study over a 20-year period: Failures in M1 Courses and Step 1. *Anatomical Sciences Education*, *8*(2), 120–125. https://doi.org/10.1002/ase.1462

Butler, A. C., & Roediger, H. L. (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology*, *19*(4–5), 514–527. https://doi.org/10.1080/09541440701326097

Charting Outcomes in the Match for U.S. Allopathic Seniors. (2016, September). Retrieved from http://www.nrmp.org/wp-content/uploads/2016/09/Charting-Outcomes-US-Allopathic-Seniors-2016.pdf

Dempster, F. N. (1989). Spacing effects and their implications for theory and practice. *Educational Psychology Review*, *1*(4), 309–330. https://doi.org/10.1007/BF01320097

Deng, F., Gluckstein, J. A., & Larsen, D. P. (2015). Student-directed retrieval practice is a predictor of medical licensing examination performance. *Perspectives on Medical Education*, *4*(6), 308–313. https://doi.org/10.1007/s40037-015-0220-x

Dobson, J. L. (2011). Effect of selected “desirable difficulty” learning strategies on the retention of physiology information. *Advances in Physiology Education*, *35*(4), 378–383. https://doi.org/10.1152/advan.00039.2011

Dobson, J. L. (2012). Effect of uniform versus expanding retrieval practice on the recall of physiology information. *Advances in Physiology Education*, *36*(1), 6–12. https://doi.org/10.1152/advan.00090.2011

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving Students’ Learning With Effective Learning Techniques: Promising Directions From Cognitive and Educational Psychology. *Psychological Science in the Public Interest*, *14*(1), 4–58. https://doi.org/10.1177/1529100612453266

Gandy, R. A., College, L., Herial, N. A., Khuder, S. A., & Metting, P. J. (2008). Use of Curricular and Extracurricular Assessments to Predict Performance on the United States Medical Licensing Examination (USMLE) Step 1: A Multi-Year Study. *Learning Assistance Review*, *13*(2), 27–35.

Gauer, J. L., & Jackson, J. B. (2017). The association of USMLE Step 1 and Step 2 CK scores with residency match specialty and location. *Medical Education Online*, *22*(1). https://doi.org/10.1080/10872981.2017.1358579

Giordano, C., Hutchinson, D., & Peppler, R. (2016). A Predictive Model for USMLE Step 1 Scores. *Cureus*, *8*(9). https://doi.org/10.7759/cureus.769

Gohara, S., Shapiro, J. I., Jacob, A. N., Gandy, R. A., Metting, P. J., Gold, J., & Kleshinski, J. (2011). Author Note: Robyn Gandy is now at Department of Psychology, Lourdes College. *Learning Assistance Review*, *16*(1), 11–20.

Graduation and Attrition Rates of U.S. Medical Students. (2018). Association of American Medical Colleges. https://www.aamc.org/download/492842/data/graduationratesandattritionratesofu.s.medicalstudents.pdf

Hattie, J. A. C., & Donoghue, G. M. (2016). Learning strategies: a synthesis and conceptual model. *Npj Science of Learning*, *1*(1), 16013. https://doi.org/10.1038/npjscilearn.2016.13

Kapur, M. (2016). Examining Productive Failure, Productive Success, Unproductive Failure, and Unproductive Success in Learning. *Educational Psychologist*, *51*(2), 289–299. https://doi.org/10.1080/00461520.2016.1155457

Kapur, M., & Bielaczyc, K. (2012). Designing for Productive Failure. *Journal of the Learning Sciences*, *21*(1), 45–83. https://doi.org/10.1080/10508406.2011.591717

Karpicke, J. D., Lehman, M., & Aue, W. R. (2014). Retrieval-Based Learning. In *Psychology of Learning and Motivation* (Vol. 61, pp. 237–284). https://doi.org/10.1016/B978-0-12-800283-4.00007-1

Kerfoot, B. P., DeWolf, W. C., Masser, B. A., Church, P. A., & Federman, D. D. (2007). Spaced education improves the retention of clinical knowledge by medical students: a

randomised controlled trial. *Medical Education*, *41*(1), 23–31. https://doi.org/10.1111/j.1365-2929.2006.02644.x

Kornell, N., & Bjork, R. A. (2008). Learning Concepts and Categories: Is Spacing the "Enemy of Induction"? *Psychological Science*, *19*(6), 585–592. https://doi.org/10.1111/j.1467-9280.2008.02127.x

Kumar, A. D., Shah, M. K., Maley, J. H., Evron, J., Gyftopoulos, A., & Miller, C. (2015). Preparing to take the USMLE Step 1: a survey on medical students' self-reported study habits. *Postgraduate Medical Journal*, *91*(1075), 257–261. https://doi.org/10.1136/postgradmedj-2014-133081

Larsen, D. P., Butler, A. C., & Roediger III, H. L. (2008). Test-enhanced learning in medical education. *Medical Education*, *42*(10), 959–966. https://doi.org/10.1111/j.1365-2923.2008.03124.x

Larsen, D. P., Butler, A. C., & Roediger III, H. L. (2009). Repeated testing improves long-term retention relative to repeated study: A randomised controlled trial. *Medical Education*, *43*(12), 1174–1181. https://doi.org/10.1111/j.1365-2923.2009.03518.x

Lesgold, A., Glaser, R., Rubinson, H., Klopfer, D., Feltovich, P., & Wang, Y. (1988). Expertise in a Complex Skill: Diagnosing X-Ray Pictures. In *The Nature of Expertise* (p. 32). Psychology Press.

Little, J. L., Bjork, E. L., Bjork, R. A., & Angello, G. (2012). Multiple-Choice Tests Exonerated, at Least of Some Charges: Fostering Test-Induced Learning and Avoiding Test-Induced

Forgetting. *Psychological Science*, *23*(11), 1337–1344. https://doi.org/10.1177/0956797612443370

Matos, J., Petri, C. R., Mukamal, K. J., & Vanka, A. (2017). Spaced education in medical residents: An electronic intervention to improve competency and retention of medical knowledge. *PLOS ONE*, *12*(7), e0181418. https://doi.org/10.1371/journal.pone.0181418

Mayer, R. E. (2008). Applying the Science of Learning: Evidence- Based Principles for the Design of Multimedia Instruction. *American Psychologist*, *63*(8), 760–769.

Mayer, R. E., & Anderson, R. B. (1992). The Instructive Animation: Helping Students Build Connections Between Words and Pictures in Multimedia Learning. *Journal of Educational Psychology*, *84*(4), 444–452.

McConnell, M. M., St-Onge, C., & Young, M. E. (2015). The benefits of testing for learning on later performance. *Advances in Health Sciences Education*, *20*(2), 305–320. https://doi.org/10.1007/s10459-014-9529-1

McDaniel, M. A., & Butler, A. C. (2010). A Contextual Framework for Understanding When Difficulties Are Desirable. In *Successful remembering and successful forgetting: A festschrift in honor of Robert A. Bjork* (pp. 175–198). Psychology Press.

McDaniel, M. A., & Donnelly, C. M. (1996). Learning With Analogy and Elaborative Interrogation. *Journal of Educational Psychology*, *88*(3), 508–519.

McDougle, L., Mavis, B. E., Jeffe, D. B., Roberts, N. K., Ephgrave, K., Hageman, H. L., … Andriole, D. A. (2013). Academic and professional career outcomes of medical school

graduates who failed USMLE Step 1 on the first attempt. *Advances in Health Sciences Education*, *18*(2), 279–289. https://doi.org/10.1007/s10459-012-9371-2

Morrison, C. A., Ross, L. P., Fogle, T., Butler, A., Miller, J., & Dillon, G. F. (2010). Relationship Between Performance on the NBME Comprehensive Basic Sciences Self-Assessment and USMLE Step 1 for U.S. and Canadian Medical School Students: *Academic Medicine*, *85*, S98–S101. https://doi.org/10.1097/ACM.0b013e3181ed3f5c

Moulton, C.-A. E., Dubrowski, A., MacRae, H., Graham, B., Grober, E., & Reznick, R. (2006). Teaching Surgical Skills: What Kind of Practice Makes Perfect?: A Randomized, Controlled Trial. *Transactions of the ... Meeting of the American Surgical Association*, *124*, 66–75. https://doi.org/10.1097/01.sla.0000234808.85789.6a

Pavlik, P. I., & Anderson, J. R. (2005). Practice and Forgetting Effects on Vocabulary Memory: An Activation-Based Model of the Spacing Effect. *Cognitive Science*, *29*(4), 559–586. https://doi.org/10.1207/s15516709cog0000_14

Results of the 2016 NRMP Program Director Survey. (2016, June). Retrieved from http://www.nrmp.org/wp-content/uploads/2016/09/NRMP-2016-Program-Director-Survey.pdf

Roediger III, H. L., Putnam, A. L., & Smith, M. A. (2011). Ten Benefits of Testing and Their Applications to Educational Practice. In *Psychology of Learning and Motivation* (Vol. 55, pp. 1–36). https://doi.org/10.1016/B978-0-12-387691-1.00001-6

Rohrer, D. (2012). Interleaving Helps Students Distinguish among Similar Concepts. *Educational Psychology Review*, *24*(3), 355–367. https://doi.org/10.1007/s10648-012-9201-3

Rohrer, D., & Pashler, H. (2007). Increasing Retention Without Increasing Study Time. *Current Directions in Psychological Science*, *16*(4), 183–186. https://doi.org/10.1111/j.1467-8721.2007.00500.x

Rohrer, D., & Pashler, H. (2010). Recent Research on Human Learning Challenges Conventional Instructional Strategies. *Educational Researcher*, *39*(5), 406–412. https://doi.org/10.3102/0013189X10374770

Schmidmaier, R., Ebersbach, R., Schiller, M., Hege, I., Holzer, M., & Fischer, M. R. (2011). Using electronic flashcards to promote learning in medical students: Retesting versus restudying: Efficacy of retesting versus restudying. *Medical Education*, *45*(11), 1101–1110. https://doi.org/10.1111/j.1365-2923.2011.04043.x

Schwartz, L. F., Lineberry, M., Park, Y. S., Kamin, C. S., & Hyderi, A. A. (2018). Development and Evaluation of a Student-Initiated Test Preparation Program for the USMLE Step 1 Examination. *Teaching and Learning in Medicine*, *30*(2), 193–201. https://doi.org/10.1080/10401334.2017.1386106

Smith, B. L., Holliday, W. G., & Austin, H. W. (2010). Students' comprehension of science textbooks using a question-based reading strategy. *Journal of Research in Science Teaching*, *47*(4), 363–379. https://doi.org/10.1002/tea.20378

Strowd, R. E., & Lambros, A. (2010). Impacting student anxiety for the USMLE Step 1 through process-oriented preparation. *Medical Education Online*, *15*(1), 4880. https://doi.org/10.3402/meo.v15i0.4880

Sumeracki, M. A., & Weinstein, Y. (2018). Six Strategies for Effective Learning: *Academic Medicine*, *93*(4), 666. https://doi.org/10.1097/ACM.0000000000002091

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive Architecture and Instructional Design: 20 Years Later. *Educational Psychology Review*, *31*(2), 261–292. https://doi.org/10.1007/s10648-019-09465-5

Thadani, R. A., Swanson, D. B., & Galbraith, R. M. (2000). A Preliminary Analysis of Different Approaches to Preparing for the USMLE Step 1: *Academic Medicine*, *75*(Supplement), S40–S42. https://doi.org/10.1097/00001888-200010001-00013

Thiede, K. W., Anderson, M. C. M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology*, *95*(1), 66–73. https://doi.org/10.1037/0022-0663.95.1.66

Thiede, K. W., Dunlosky, J., Griffin, T. D., & Wiley, J. (2005). Understanding the Delayed-Keyword Effect on Metacomprehension Accuracy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(6), 1267–1280. https://doi.org/10.1037/0278-7393.31.6.1267

Van Dongen, K. W., Mitra, P. J., Schijven, M. P., & Broeders, I. A. M. J. (2011). Distributed versus massed training: Efficiency of training psychomotor skills. *Surgical Techniques Development*, *1*(1), 17. https://doi.org/10.4081/std.2011.e17

Weinstein, Y., Madan, C. R., & Sumeracki, M. A. (2018). Teaching the science of learning. *Cognitive Research: Principles and Implications*, *3*(1), 2. https://doi.org/10.1186/s41235-017-0087-y

Werner, L. S., & Bull, B. S. (2003). The effect of three commercial coaching courses on Step One USMLE performance. *Medical Education*, *37*(6), 527–531. https://doi.org/10.1046/j.1365-2923.2003.01534.x

West, C., Kurz, T., Smith, S., & Graham, L. (2014). Are study strategies related to medical licensing exam performance? *International Journal of Medical Education*, *5*, 199–204. https://doi.org/10.5116/ijme.5439.6491

Yan, V. X., Clark, C. M., & Bjork, R. A. (2016). Memory and Metamemory Considerations in the Instruction of Human Beings Revisited: Implications for Optimizing Online Learning. In *From the Laboratory to the Classroom: Translating Science of Learning for Teachers*. Routledge.

Zhang, C., Rauchwarger, A., Toth, C., & O'Connell, M. (2004). Student USMLE Step 1 Preparation and Performance. *Advances in Health Sciences Education*, *9*, 291–2